On Combining Alignment Techniques

Tordai, A.

2012

document version
Publisher's PDF, also known as Version of record

Summary

Cultural collections are composed (of images) of cultural heritage objects, their description in metadata, and associated vocabularies with terms used in the metadata. Collections vary not only in content, but as collection metadata was created independently by each institution, they also employ a variety of formats making the integration of several collections into a single virtual collection difficult. The metadata and vocabularies first need to be converted to a common format and then linked to each other.

In this thesis we start by investigating the steps necessary for the integration of a collection into a large virtual collection. We identify four distinct steps and illustrate them in a case study. First, we convert all vocabularies accompanying the collection metadata into a common format (SKOS). Second, we convert the metadata schema, which is the set of elements describing the metadata, to more generic schemas (Dublin Core and VRA). Third, we convert the values of the metadata. In this step we identify meaningful values that either come from vocabularies or may be identifiable objects in their own right, such as a painter whose works are contained in several collections. As a last option, we keep values as text. In the fourth and last step we align collection vocabularies to vocabularies that are already part of the virtual collection, by creating mappings between similar concepts. As a result, the vocabularies that form the indexes of the collections also become integrated. The vocabularies may be aligned manually, which is a time-intensive task, or automatically. There is a multitude of alignment tools available that implement one or more alignment techniques. Although these tools are evaluated and compared in the Ontology Alignment Evaluation Initiative (OAEI), there is no methodological advice on how to align vocabularies from start to finish.

In the remainder of the thesis we focus on vocabulary alignment. Our main research question is as follows: *How can we combine vocabulary alignment techniques, assess their performance and evaluate alignments?*

In Chapters 3 and 4 we study how we can combine vocabulary alignment techniques and assess their performance. In our experiment in Chapter 3, two vocabularies are aligned using string matching and off-the-shelf alignment tools. We find that although the individual tools perform relatively poorly, by selecting subsets of mappings we are able to boost the quality of the mappings in comparison to the original set.

In Chapter 4 we perform a similar experiment on large vocabularies. Due to the size of the vocabularies we were unable to use the off-the-shelf tools from Chapter 3 and resort to simple string matching techniques to generate mappings. As the number of generated mappings was too high to evaluate in their entirety, we took random samples in order to be able to assess the value of the alignment techniques and compare them. In a step-by-step method we select subsets of mappings of higher quality than the original set of mappings.

In the Chapter 5 we focus on reusing existing equivalence mappings between multiple vocabularies and generate composed mappings. We investigate whether composed mappings are transitive by setting up experiments with vocabularies from different domains: the medicine and cultural heritage library domain. We generate composed mappings and examine the number of composed mappings and their quality in comparison to the mappings used to create them. Our findings matched our expectations about mapping composition. For example, we found that vocabularies that overlap in the domain they describe are more likely to have high quality composed mappings. One aspect that deserves more study is the fact that the quality of composed mappings was higher than expected, almost as high as the quality of mappings used to create the composed mappings.

In order to evaluate the quality of an alignment at least part of the mappings need to be evaluated manually. For manual evaluation to be reliable, at least a subset of mappings needs to be evaluated by independent human raters. Then their agreement level needs to be calculated, and if it is sufficiently high the result of the manual evaluation can be used as a test set or reference alignment. In our experiments we found that the agreement level was in some cases rather low. In Chapters 6 and 7 we study the level of agreement between multiple raters in several manual evaluation experiments, and attempt to identify the causes for low agreement levels. We find that improved guidelines with clear descriptions of the task and evaluation categories improves the level of agreement between raters. Nevertheless there are also aspects that make evaluation difficult, such as the inherent ambiguity of concepts and their descriptions, and the background knowledge of raters which influences their interpretation of concepts. Additionally, characteristics of vocabularies, such as their representation and their domain and the selection of suitable matching categories also influence the outcome of an evaluation.

In this thesis we show that combining alignment techniques through an interactive process is an effective and transparent method for generating high quality mappings. We also show that a large number of factors influence manual evaluation. These factors include the characteristics of vocabularies, the matching categories and the inherent vagueness of concepts.

We conclude with a method for evaluating alignments which includes reporting agreement levels between raters, reporting on how disagreements were dealt with, and emphasizes that the evaluation needs to be made public. Second, we propose a method for aligning vocabularies, where we keep track of the source of mappings. As the alignment is performed in explicit steps it is possible to select mappings of a specific quality, and thus the alignment can be tailored to any kind of application.